DOCUMENT RESUME

ED 106 336 TM 004 451

AUTHOR Hubert, Lawrence; Schultz, James
TITLE Hierarchical Clustering and the Concept of Space Distortion.
PUB DATE [Apr 75]
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Assciation (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Behavioral Sciences; *Classification; *Cluster Analysis; *Cluster Grouping; Data Analysis; Matrices; *Space

ABSTRACT

An empirical assesssment of the space distortion properties of two prototypic hierarchical clustering procedures is given in terms of an occupancy model developed from combinatorics. Using one simple example, the single-link and complete-link clustering strategies now in common use in the behavioral sciences are empirically shown to be space contracting and space dilating, respectively. An extension to an intermediate r-diameter clustering criterion defined in between the complete-link and single-link extremes is discussed briefly along with a technique for preprocessing the original proximity measures to make the results obtained from these two extreme clustering procedures more consistent with the data. (Author)

ED106336

HIERARCHICAL CLUSTERING AND THE CONCEPT OF SPACE DISTORTION

Lawrence Hubert
James Schultz*

The University of Wisconsin, Madison

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 004 451

HIERARCHICAL CLUSTERING AND THE CONCEPT OF SPACE DISTORTION

Abstract

An empirical assessment of the space distortion properties of two prototypic hierarchical clustering procedures is given in terms of an occupancy model developed from combinatorics. Using one simple example, the single-link and complete-link clustering strategies now in common use in the behavioral sciences are empirically shown to be space contracting and space dilating, respectively. An extension to an intermediate r-diameter clustering criterion defined in between the complete-link and single-link extremes is discussed briefly along with a technique for preprocessing the original proximity measures to make the results obtained from these two extreme clustering procedures more consistent with the data.

HIERARCHICAL CLUSTERING AND THE CONCEPT OF SPACE DISTORTION

1. INTRODUCTION

One of the basic properties that supposedly characterizes all hierarchical clustering methods can be defined in terms of a particular strategy's ability to conserve "space" in the region surrounding a new subset formed during the clustering process. Although the necessary concepts introduced by Lance and Williams (1967) are obviously vague and lack formal or precise definitions of any sort, the basic idea of a space conserving or space distorting clustering procedure does seem to have substantial intuitive appeal, especially for an applied researcher well-versed in the various techniques for partitioning a set of objects into homogeneous subgroups. The intent of this paper is to concentrate on two so-called space distorting clustering strategies now in general use in the behavioral sciences, commonly called the single-link and the complete-link methods. The discussion below offers one possible combinatorial framework for characterizing a space distorting strategy, and in particular, for the terms "space contracting" and "space dilating" used in connection with the prototypic single-link and complete-link clustering schemes, respectively.

2. CLUSTERING PROCEDURES

To give a brief background summary on clustering, suppose $s(\cdot,\cdot)$ is a positive real-valued symmetric function on S x S, where S is a set of n objects $\{o_1,\ldots,o_n\}$; intuitively, the proximity function $s(\cdot,\cdot)$ assigns smaller numerical values to the more similar pairs of objects, where it is assumed that $s(o_i,o_j) = 0$ if $o_i = o_j$. Both the single-link and the complete-link clustering procedures construct a sequence of partitions of S, $(\ell_o,\ldots,\ell_{n-1})$, in which ℓ_o is the trivial partition containing each object in a separate class, ℓ_{n-1} includes all objects in a single class, and ℓ_{k+1} is obtained from ℓ_k by uniting two of the subsets in ℓ_k. In particular, the complete-link hierarchy is defined by uniting those two object classes in ℓ_k with the minimum resulting diameter, where the term "diameter" is defined as the maximum of $s(\cdot,\cdot)$ over all object pairs within a subset. The single-link method, on the other hand, employs an exact opposite criterion and unites those two subsets in ℓ_k that minimize $s(\cdot,\cdot)$ over all object pairs in which each member of the pair belongs to only one of the two subsets to be joined.

A number of empirical evaluations of the single-link criterion have documented a chaining characteristic of the method, in the sense that new partitions in the sequence tend to be formed by uniting single elements with pre-existing subsets (for example, see Wishart, 1969). In other words, at a particular partition level the subsets are generally large and as a new subset is formed, it appears to move nearer to some or to all of the remaining members of S, and consequently, the region around the subset appears to contract. For the complete-link procedure, however, the subsets within a partition tend to be more equal in size,

and as new subsets are formed, they appear to move further away from the remaining subsets in the partition. Thus, the region around a newly constructed subset is said to dilate and new members of a partition are formed by uniting the smaller subsets within the previous level.

The concepts of "region," "space," and so forth used in the above discussion may be intuitively meaningful but fail to convey very precise properties. Consequently, as an alternative approach given below, the emphasis will be on operationalizing the space-distortion ideas through the number and type of subsets formed at a specific partition level by the two clustering techniques. In particular, a space contracting strategy should produce, on the average, a large subset and many small ones at each partition level, and attempt to minimize the number of subsets containing more than a single object. Conversely, a space dilating technique should construct subsets of about the same size and maximize the number of subsets containing more than one object at each partition level.

3. ASSESSING SPACE DISTORTION

As mentioned above, one of the ways of evaluating the space conserving or space distorting properties of a clustering method is in terms of the type of subsets generally produced at a specific partition level in the hierarchy. Such an evaluation, however, requires some fixed set of functions $s(\cdot,\cdot)$ that will be used for developing a frame of reference to make the phrase "generally produce" more precise. For example, since both the single-link and the complete-link procedures depend solely upon the rank order of the object pairs specified by the proximity measures, the set of functions $s(\cdot,\cdot)$ that will be employed here corresponds to the set of all assignments of the distinct integers from 1 to $n(n-1)/2$ to a fixed ordering of the $n(n-1)/2$ object pairs. Thus, for all $[n(n-1)/2]!$ possible assignments of rank, the complete-link and the single-link strategies may be used to obtain the distributions of subset types produced at the various partition levels. To carry the discussion one step further, these distributions can then be compared to an expected distribution that would be obtained if the clustering procedure were space conserving; in fact, the differences between the obtained distributions and this standard could be incorporated into some formal index of the space distortion propensity of the clustering method.

What is still lacking in this presentation is any specification of what the standard distribution should be at a particular partition level. Within the rank order framework, however, there is an obvious suggestion that can be developed from the enormous literature on occupancy problems. To be more precise, suppose there are n objects in S and the partition level under consideration consists of m subsets, i.e., the partition ℓ_{n-m}

at level n-m in the sequence $(\ell_0,\ldots,\ell_{n-1})$. The number of different ways in which the n distinguishable objects can be distributed into the m indistinguishable cells of the partition is S(n,m), a Stirling number of the second kind (Riordan, 1958). Consequently, if we apply the occupancy model as a standard for a space-conserving procedure and use our set of all possible [n(n-1)/2]! assignments of integer ranks to the object pairs, a space conserving clustering strategy should produce partitions at level n-m of the form $\{n_1,\ldots,n_m\}$ according to the distribution

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n_m}{n_m} \Big/ (bS(n,m)),$$

where $n_i \geq 1$ is the number of objects in the i^{th} subset, $1 \leq i \leq m$; $\sum_i n_i = n$, and $b = \prod_{t\in T} t!$. In this context, T is a set of integers containing the number of subsets of size n_i, $1 \leq i \leq m$, but which does not count a set of tied subsets more than once. As an example that will be carried out in more detail below, suppose n = 12 and m = 3. Then, S(12,3) = 86,526 and the probability of obtaining the partition {6,3,3} is

$$\binom{12}{6}\binom{6}{3}\binom{3}{3} \Big/ (2!86,526) = .107.$$

To make the computational burden somewhat more manageable, the following discussion is limited to the case of n = 12 and m = 3. Furthermore, instead of studying all [12·11/2]! possible permutations of the integer ranks for the object pairs, a sample of 1000 permutations was selected randomly and with replacement, and for each such assignment the complete-link and single-link partitions obtained at level 9, i.e.,

when m = 3. Table 1 presents the theoretical distribution over the partition types calculated from the general expression given above, and Table 2 presents the empirical distributions obtained from the two clustering procedures using the 1000 randomly chosen permutations of the integer ranks. Table 3 compares the empirical and the theoretical distributions obtained by collapsing over partitions in which the largest partition class has the same number of objects.

Tables 1, 2, and 3 here

Compared to the reference distribution generated under the occupancy model, Tables 2 and 3 clearly indicate the space distortion propensities of the two clustering strategies under study. As an illustration, if n and m are fixed, the random variable $X_{n,m}$, defined as the size of the largest partition class, has an expected value of 5.70 under the occupancy model. Since the sample means corresponding to the 1000 observations on $X_{n,m}$ for the complete-link scheme is 5.55 and 9.06 for the complete-link, the single-link procedure is obviously biased towards the construction of larger partition classes whereas the complete-link criterion tends to over-equalize the distribution of objects into the m subsets of the decomposition. The bias in the single-link method is very large, and rather surprisingly, the bias in the complete-link method is slight although still evident.

4. r-DIAMETER HIERARCHICAL CLUSTERING

Although many suggestions have been made for relaxing the partitioning criterion used in the complete-link method, or conversely, strengthening the single-link criterion, only a very few of these proposals maintain a sole dependence upon the rank order of the proximity values. Two of the most promising of these new approaches that, in addition, can still construct disjoint partition classes at each level of the hierarchy are due to Ling (1973) and Matula (1972). Ling uses the graph-theoretic concept of minimum degree and Matula emphasizes k-line connectivity and graph cohesiveness. The interested reader is referred to Hubert (1974b) for an extensive review of the literature relating to these techniques as well as for a more general background for the method suggested below.

As an alternative characterization of complete-link clustering, suppose the partition ℓ_k has been obtained and the partition ℓ_{k+1} is to be formed by uniting the two subsets in ℓ_k with the minimum resultant diameter. As mentioned above, the diameter of a subset of S is defined as the maximum proximity measure over all object pairs within the subset. As a convenience in defining the generalizations below, this particular concept of a diameter will be referred to as 1-diameter (see Hubert, 1974b). The concept of an r-diameter subset follows naturally from Luce's (1950) notion of a generalized clique and Peay's (1970) application of this idea to clustering by means of hierarchical overlapping subsets. In particular, the r-diameter of a subset S' of S is defined as the minimum proximity measure Q over all object pairs in S' that satisfies the following property:

for any two objects o_{i_0} and $o_{i_t} \in S'$, there is a sequence of not necessarily distinct objects $o_{i_0}, o_{i_1}, \ldots, o_{i_{r-1}}, o_{i_t}$ such that $o_{i_j} \in S'$, $0 \leq j \leq t$, and $s(o_{i_{j'}}, o_{j_{j'+1}}) \leq Q$, $0 \leq j' \leq t-1$.

Clearly, r-diameter hierarchical clustering can be defined by uniting those two subsets in ℓ_k to form ℓ_{k+1} that have a minimum resulting r-diameter. What is particularly interesting about r-diameter clustering, however, is the natural relation this generalization has to the single-link and the complete-link methods; specifically, r-diameter hierarchical clustering depends only upon the rank order of the proximity measures, produces a partition at each level of the hierarchy, and corresponds to complete-link clustering for r = 1 and to single-link clustering for r = n - 1.

From a computational point of view, the ability to perform any r-diameter clustering requires only a complete-link algorithm and a set of modified initial proximity values. More precisely, an r-diameter partition hierarchy is a complete-link partition hierarchy using a transformed proximity function

$$s^{(k)}(o_i, o_j),$$

defined as the (i,j) element in a matrix $M^{(k)}$; the matrix $M^{(k)}$ is constructed recursively in the standard manner, i.e., $M^{(k)} = M^{(k-1)} \oplus M$, where the operation $\oplus$ is specified as follows:

suppose $A_{nxn} = \{a_{ij}\}$, $B_{nxn} = \{b_{ij}\}$, then

$$A \oplus B = \{c_{ij}\}_{nxn} = \min_k \{\max\{a_{ik}, b_{kj}\}\}.$$

A similar matrix relationship is used by Peay (1970) to define a generalization of a hierarchical clique detection program using Luce's (1950) concept of a generalized clique.

Returning to the previous discussion of space-distorting clustering methods, the notion of an r-diameter subset provides a tentative compromise. The complete-link (1-diameter) and single-link (n-1 diameter) criteria can be viewed as forming the two extremes of a space distortion "continuum"; any r-diameter clustering procedure lies at some point along this continuum, and as r increases, the strategies become less space dilating and more space contracting. Consequently, it is up to the applied researcher to choose that particular point along the continuum that satisfies his specific data analytic requirements; this vagueness is necessary since no value of r will provide the general space conserving procedure for all partition levels and for all values of n.

As an example of the decision a user faces in applying the r-diameter concept, some famous published data from Rao (1952) on the similarity of 12 Indian Castes was clustered using all values of r. Hierarchies different from those obtained with the single-link and complete-link methods were constructed for values of r from 2 to 4; the single-link hierarchy was consistently produced for $r \geq 5$, and the complete-link hierarchy for $r \leq 1$. Table 4 presents the various partitions obtained at level 7, which corresponds to the level of Rao's "optimal" partition also obtained with the 1-diameter (complete-link) strategy. The progression as r increases from fairly uniform sized subsets to a single elementary subset is obvious.

Table 4 here

In general, any application of hierarchical clustering will produce some value of r, say r_0, such that for any value of r greater than or equal to r_0, r-diameter hierarchical clustering will construct the single-link hierarchy. Similarly, there will be some value of r, say r_1, such such that when r is less than or equal to r_1, the complete-link hierarchy is obtained. The size of the difference r_0-r_1 can be considered a "measure" of how divergent the results obtained from the complete-link and single-link hierarchical clustering methods will be when used on the same set of data. In fact, using the paradigm discussed by Hubert (1974a), the sample distribution of r_0-r_1 could be approximated for each n under a hypothesis of randomness in the assignment of proximity ranks to the object pairs. Although these distributions could be of value is assessing the "randomness" of the basic set of proximity measures, they would be somewhat expensive to obtain and the topic is not persued here.

5. THE "SMOOTHING" OF PROXIMITY VALUES

One of the basic difficulties facing the user of hierarchical clustering is in the need to make a final selection of one clustering result from the many that are easily available. Within the behavioral sciences, the single-link method usually is ignored as a viable strategy since the solutions obtained by the method tend to be substantively less interpretable than those produced by the complete-link criteria (for instance, consult Fillenbaum & Rapoport, 1971, or Johnson, 1967). The biological sciences, on the other hand, favor the single-link strategy over the complete-link and apparently obtain single-link clustering solutions that are usually more amenable to substantive evaluation; some controversy does exist, however (see Jardine & Sibson, 1971).

Possibly, one of the reasons for the divergence of empirical opinion in the two disciplines regarding the utility of single-link clustering may result from the different types of data usually analyzed. The biological sciences concentrate on proximity values that have metric properties by defining proximity through appropriate metric measures based upon numerical attributes that characterize the objects to be clustered. Many of the applications in the behavioral sciences, however, are based upon data that fail to satisfy the standard triangle inequality, for example, proximity values are commonly produced by direct estimation procedures from a group of subjects (see Shepard, 1972). If it is true that the non-metric properties of the standard behavioral science data will generally prevent reasonable hierarchies from being constructed by means of the single-link method, alternative single-link related strategies that have rather elegant axiomatic justifications (Jardine & Sibson, 1971)

may be developed by requiring an initial redefinition of proximity to make the input data originally more "metric." In fact, the preprocessing of proximity measures may be beneficial in obtaining meaningful results from other clustering strategies, or for that matter, from alternative data analytic techniques such as multidimensional scaling. This initial preprocessing will generally require more than the ordinal information in the original proximity measures, but at least in the suggestion given below, any strong interval scale dependencies are avoided; also, if the original data is already metric, then no change is effected by the preprocessing.

As one possible suggestion for transforming the original matrix of proximity measures, a simple modification of the equation used to preprocess the proximity data for the r-diameter clustering could be used. Specifically, instead of the basic matrix $M = P^{(1)}$, we could input either $P^{(2)}, \ldots,$ or $P^{(n-1)}$ to a clustering strategy, where

$$P^{(k)} = P^{(k-1)} \otimes P^{(1)}, \text{ and}$$

$$\{a_{ij}\} \otimes \{b_{ij}\} = \{c_{ij}\} = \min_{k}\{a_{ik} + b_{kj}\}.$$

Essentially, the operation $\otimes$ forces the triangle inequality to hold in stages; the final matrix $P^{(n-1)}$ has to be metric and contains the shortest "distance" between any two objects o_i and o_j in a complete graph with edges weighted by the proximities in M. These matrix operations are well-known in graph theory and are discussed fully in Marshall (1971).

If the single-link or complete-link clustering procedures are applied to $P^{(k)} \neq M$, then the resultant partition depends upon more than the ordinal information in the matrix M. The case of $P^{(2)}$ is particularly

instructive since the dependence on M has been given the specific name of "hypermonotone invariant" by Hubert (1972). More precisely, any monotone increasing transformation of the proximity measures in M is called hypermonotone invariant if the transformation also preserves the ordering of all first differences between the measures (Suppes & Zines, 1963). Thus, any clustering procedure that relies only upon the rank order of the basic input measures will become a hypermonotone invariant procedure with respect to M when applied to $P^{(2)}$. The reader is referred to Hubert (1972) for a discussion of the value of this type of dependence or to Suppes and Zinnes (1963) and Kendall (1962) for a discussion of the basic framework that would allow a similar invariance discussion for $P^{(k)}$, when $k > 2$.

As a small and exploratory example of the effect of using the matrices $P^{(k)}$ for $k \geq 2$ instead of the original matrix $P^{(1)} = M$, the proximity values between Rao's 12 Indian castes were reallocated to the 66 object pairs at random. This procedure was replicated 10 times, and for each such reassignment, the single-link and complete-link partition hierarchies were obtained using the input matrices $P^{(k)}$, $1 \leq k \leq n-1$. Using the data obtained in this manner, Table 5 gives a goodness-of-fit measure between each constructed hierarchy and the input data for selected values of k. Specifically, Hubert (1974a) defines the adequacy of a particular partition hierarchy through a rank order correlation index γ; the index is obtained between the object pairs ranked on the basis of the input proximity values and a reranking of the object pairs produced by the partition hierarchy itself. This latter reranking is defined by assigning to each object pair the partition level at which the object first appears within the same subset.

Table 5 here

The information given in Table 5 leads to several rather interesting empirical conjectures: first of all, the hypothesis of randomness in assigning proximity ranks to the object pairs used in generating the sampling distributions given by Hubert, appears to be a "lenient" hypothesis. As k increases, the mean γ values obtained for $P^{(k)}$ increase; consequently, as the data become "more metric," and even though they are still based upon an initial random assignment of the proximity measures in $P^{(1)}$, it becomes much easier to reject the basic randomness hypothesis using the sampling distributions generated from $P^{(1)}$. Thus a rejection of randomness in the assignment of proximity values by a sizable γ value appears to be a rather weak criterion that a clustering result should easily pass; if rejection is not possible, then any further substantive interpretation of the partition hierarchy is of dubious worth. Secondly, as the initial data become "more metric," the single-link strategy appears to produce a more adequate partition hierarchy, at least in terms of γ, and the increase in γ is more noticeable for the single-link procedure than for the complete-link. In fact, for the "metric" data sets, i.e., $P^{(11)}$, the mean γ values over the 10 replications are fairly close for the two clustering schemes; clearly, the single-link procedure S is affected more by the non-metric nature of the original data. This result suggests that the inability of the single-link method to elicit substantively meaningful structures in the behavioral sciences may be due in part to the non-metric nature of our common proximity measures. Obviously, the chaining characteristics of single-link clustering are still present (see Wishart, 1969), but apparently, there is some mitigation of the problem at least as measured by the overall γ index.

REFERENCES

Fillenbaum, S. and Rapoport, A. (1971). Structures in the Subjective Lexicon. New York: Academic Press.

Hubert, L. (1972). Some extensions of Johnson's hierarchical clustering algorithms. Psychometrika, 37, 261-74.

Hubert, L. (1974a). Approximate evaluation techniques for the single-link and complete-link clustering procedures. J. Amer. statist. Assoc. 69, 698-704.

Hubert, L. (1974b). Some applications of graph theory to clustering. Psychometrika, 39, 283-309.

Jardine, N. and Sibson, R. (1971). Mathematical Taxonomy. New York: Wiley.

Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32, 241-54.

Kendall, M. G. (1962). Ranks and measures. Biometrika, 49, 133-7.

Lance, G. N. and Williams, W. T. (1967). A general theory of classifactory sorting strategies I. Hierarchical systems. Comp. J. 10, 373-80.

Ling, R. F. (1973). A probability theory of cluster analysis. J. Amer. statist. Assoc. 68, 159-164.

Luce, R. D. (1950). Connectivity and generalized cliques in sociometric group structure. Psychometrika, 15, 169-90.

Marshall, C. W. (1971). Applied Graph Theory. New York: Wiley.

Matula, D. W. (1972). k-components, clusters, and slicings in graphs. SIAM J. Appl. Math. 22, 459-80.

Peay, E. R. (1970). Nonmetric Grouping: Clusters and Cliques. Michigan Mathematical Psychology Program, 70-5, Ann Arbor, Michigan.

Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. New York: Wiley.

Riordan, J. (1958). An Introduction to Combinatorial Analysis. New York: Wiley.

Shepard, R. N. (1972). A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, and S. B. Nerlove (Eds.), Multidimensional Scaling - Volume I. New York: Seminar Press.

Suppes, P. and Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. Bush, and E. Galanter (Eds.), Handbook of Mathematical Psychology - Volume I. New York: Wiley.

Wishart, D. (1964). Mode analysis: A generalization of nearest neighbor which reduces chaining effects. In A. J. Cole (Ed.), Numerical Taxonomy. New York: Academic Press.

FOOTNOTE

*Equal authorship is implied.

TABLE 1. THEORETICAL DISTRIBUTION OVER PARTITION TYPES AT LEVEL 9 FOR n = 12 - OCCUPANCY MODEL

Partition Type	Number of Realizations	Probability
{10,1,1}	66	.001
{9,2,1}	660	.008
{8,3,1}	1980	.023
{8,2,2}	1485	.017
{7,3,2}	7920	.092
{7,4,1}	3960	.046
{6,5,1}	5544	.064
{6,4,2}	13860	.160
{6,3,3}	9240	.107
{5,5,2}	8316	.096
{5,4,3}	27720	.320
{4,4,4}	5775	.067

TABLE 2. EMPIRICAL DISTRIBUTIONS OVER PARTITION TYPES AT LEVEL 9 FOR n = 12 USING THE SINGLE-LINK AND COMPLETE-LINK METHODS--SAMPLE SIZE OF 1000

Partition Types: Single-link	Partition Types: Complete-link {10,1,1}	{9,2,1}	{8,3,1}	{8,2,2}	{7,3,2}	{7,4,1}	{6,5,1}	{6,4,2}	{6,3,3}	{5,5,2}	{5,4,3}	{4,4,4}	Proportions
{10,1,1}	0	3	4	15	45	5	13	103	58	64	193	43	.546
{9,2,1}	0	0	0	4	21	1	3	48	24	28	66	14	.209
{8,3,1}	0	0	1	2	11	0	3	16	6	4	26	12	.081
{8,2,2}	0	0	0	0	1	0	0	5	1	5	10	0	.022
{7,3,2}	0	0	0	1	3	0	0	5	0	3	8	1	.021
{7,4,1}	0	0	0	0	2	0	0	14	5	5	11	3	.040
{6,5,1}	0	0	1	0	7	0	1	3	5	3	24	2	.046
{6,4,2}	0	0	0	1	3	0	0	8	0	1	2	2	.017
{6,3,3}	0	0	0	0	1	0	0	1	1	0	0	0	.003
{5,5,2}	0	0	0	0	1	0	0	1	1	1	2	1	.007
{5,4,3}	0	0	0	0	0	0	0	2	1	1	1	3	.008
{4,4,4}	0	0	0	0	0	0	0	0	0	0	0	0	.000
Proportions	.000	.003	.006	.023	.095	.006	.020	.206	.102	.115	.343	.081	

TABLE 3. COMPARISON OF EMPIRICAL AND THEORETICAL DISTRIBUTIONS OVER PARTITION TYPES AT LEVEL 9 FOR n = 12 - COLLAPSED OVER PARTITION TYPES CONTAINING THE SAME SIZE LARGEST SUBSET

Partition Type	Theoretical Probability	Empirical Probabilities	
		Single-link	Complete-link
{10,1,1}	.001	.546	.000
{9,2,2}	.008	.209	.003
{8,3,1} {8,2,2}	.040	.103	.029
{7,4,1} {7,3,2}	.137	.061	.101
{6,5,1} {6,4,2} {6,3,3}	.331	.066	.328
{5,5,2} {5,4,3} {4,4,4}	.416	.015	.458

TABLE 4. LEVEL 7 PARTITIONS OBTAINED FROM RAO'S CASTE DATA USING r-DIAMETER CLUSTERING - ABBREVIATIONS ARE RAO'S

r	Partition
1	$\{\{A_1,A_2,A_3,A_4\},\{B_1,B_2\},\{Ch,M\},\{Bh,D\},\{C_1,C_2\}\}$
2	$\{\{A_1,A_2,A_3,A_4\},\{B_1,B_2,C_2\},\{Ch,M\},\{Bh,D\},\{C_1\}\}$
3	$\{\{B_1,B_2,A_1,A_2,A_3,A_4\},\{Ch,M\},\{Bh,D\},\{C_1\},\{C_2\}\}$
4	$\{\{B_1,B_2,A_1,A_2,A_3,A_4\},\{Ch,M\}\ \{Bh,D\},\{C_1\},\{C_2\}\}$
5	$\{\{B_1,B_2,A_1,A_2,A_3,A_4,Ch,M\}\ \{Bh\},\{D\},\{C_1\},\{C_2\}\}$

TABLE 5. SELECTED GOODNESS-OF-FIT γ VALUES FOR SINGLE-LINK AND COMPLETE-LINK CLUSTERING USING THE MATRICES $P^{(k)}$. SAMPLE SIZE IS 10

Data Set		Single-link				Complete-link			
	k:	1	2	4	11	1	2	4	11
1		.13	.39	.59	.59	.39	.39	.70	.70
2		.27	.53	.65	.65	.39	.56	.39	.39
3		.20	.49	.72	.73	.36	.51	.68	.69
4		.31	.53	.60	.60	.42	.64	.44	.44
5		.25	.46	.60	.60	.49	.60	.66	.66
6		.26	.53	.59	.59	.38	.64	.54	.54
7		.13	.54	.73	.73	.33	.44	.75	.75
8		.13	.53	.62	.62	.40	.64	.68	.68
9		.41	.72	.82	.82	.40	.51	.78	.78
10		.47	.80	.85	.85	.29	.77	.85	.85
	Means:	.26	.55	.68	.68	.39	.57	.65	.65
Original Data:		.50	.62	.64	.64	.48	.65	.68	.68